COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–36

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12362

Processing Conversational Implicatures: Alternatives and Counterfactual Reasoning*

Bob van Tiel,[a] Walter Schaeken[b]

[a]*Department of Languages and Literature, Université Libre de Bruxelles*
[b]*Laboratory for Experimental Psychology, University of Leuven*

Received 9 December 2014; received in revised form 24 November 2015; accepted 15 December 2015

Abstract

In a series of experiments, Bott and Noveck (2004) found that the computation of scalar inferences, a variety of conversational implicature, caused a delay in response times. In order to determine what aspect of the inferential process that underlies scalar inferences caused this delay, we extended their paradigm to three other kinds of inferences: free choice inferences, conditional perfection, and exhaustivity in "it"-clefts. In contrast to scalar inferences, the computation of these three kinds of inferences facilitated response times. Following a suggestion made by Chemla and Bott (2014), we propose that the time it takes to compute a conversational implicature depends on the structural characteristics of the required alternatives.

Keywords: Pragmatics; Alternatives; Scalar inference; Conversational implicature; Sentence processing; Conditional perfection; Free choice inferences; Clefts

1. Introduction

1.1. Conversational implicatures and utterance processing

In his *Logic and conversation*, Grice introduced conversational implicatures as propositions that can be calculated on the basis of the literal meaning of an utterance and the assumption that the speaker is cooperative. To illustrate, suppose a student tells you, "I did some of my homework." The literal interpretation of this utterance is that she did at

Correspondence should be sent to Bob van Tiel, Université Libre de Bruxelles, Avenue Franklin Roosevelt 50, Bruxelles 1050, Belgium. E-mail: bobvantiel@gmail.com

*This work originates in the Ph.D. thesis of the first author (van Tiel, 2014b, chapter 3.3). An improved version of that chapter was originally submitted to *Cognitive Science*. Based on the comments made by the reviewers, we decided to reconstruct and redo the experiment. A brief discussion of the original experiment can be found in Data S1 in the Supplementary Material.

least some and possibly all of her homework. Nevertheless, you are likely to infer that she did not do all of her homework. This *scalar inference* is often explained as a conversational implicature along the following lines: The speaker could have been more informative and hence cooperative by using the alternative "I did all of my homework." Why didn't she? Presumably because she did not do all of her homework (cf. Horn, 1972 2004).

Grice's definition was intended as a criterion for demarcating different kinds of meaning rather than a description of the cognitive processes that underlie the derivation of conversational implicatures in actual conversations, as he emphasizes in the following passage:

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replacable by an argument, the implicature (if present at all) will not count as a CONVERSATIONAL implicature [. . .]. (p. 50)

Despite his careful formulation, Grice is often credited with the view that listeners actively calculate conversational implicatures (e.g., Levinson, 2000; Millikan, 2005; Noveck & Sperber, 2007; Origgi & Sperber, 2000; Récanati, 1995; Sperber & Wilson, 1987; Wilson, 2000). One of the corollaries of this view is that listeners access the literal interpretation of an utterance before calculating conversational implicatures. After all, the literal interpretation is a premise in the reasoning process that leads to a conversational implicature. This corollary is sometimes referred to as the *literal-first hypothesis* (e.g., Degen & Tanenhaus, 2015b).

A number of theorists have rejected the literal-first hypothesis attributed to Grice. For example, Récanati (1995) argues that the literal interpretation and the interpretation enriched with a conversational implicature are processed in parallel. Upon encountering a potentially ambiguous constituent (e.g., "some"), both its literal ("at least some and possibly all") and its enriched ("at least some but not all") meanings become activated and enter into the interpretative process. The meaning that results in an interpretation that is supported by the broader context is retained while the other meaning is suppressed. According to this approach, it is possible for an utterance to immediately receive an enriched interpretation without first establishing the literal interpretation, as long as it is supported by the context (see, e.g., Chierchia, 2006; Chierchia, Fox, & Spector, 2012; Geurts, 2010; Levinson, 2000; Pickering & Garrod, 2004; Sperber & Wilson, 1987, for similar proposals).

Both sides of the debate have emphasized the importance of psycholinguistic evidence to decide whether the literal-first hypothesis is correct. To this end, researchers have analyzed the processing times for various kinds of inferences that are commonly explained as conversational implicatures, such as indirect questions (Clark, 1979; Gibbs, 1983), sarcasm (Gibbs, 1986a), idioms (Gibbs, 1986b; Ortony, Schallert, Reynolds, & Antos, 1978), metaphors (Gildea & Glucksberg, 1983; Glucksberg, Gildea, & Bookin, 1982; Harris, 1976), and hyperbole (Deamer, Pouscoulous, & Breheny, 2010). For all of these kinds of pragmatic inferences, it was found that the interpretation enriched with a conversational implicature was processed at least as quickly as the literal interpretation, thus casting doubt on the literal-first hypothesis.

There is, however, one class of information commonly explained as a conversational implicature that seems to confirm the literal-first hypothesis, namely the aforementioned scalar inferences. In their seminal paper, Bott and Noveck (2004) asked participants to indicate the truth value of sentences like:

(1) a. Some dogs are mammals.
 b. Some parrots are birds.

Such sentences are true if "some" receives its literal interpretation as "at least some and possibly all," but false if a scalar inference is computed and "some" is interpreted as excluding "all." In line with the literal-first hypothesis, it was found that "false" answers took significantly longer than "true" answers.

The state of the art thus leaves us with the following conundrum: Why is the derivation of scalar inferences, unlike many other types of conversational implicature, associated with a processing cost? A solution to this problem is paramount in determining whether the literal-first hypothesis is correct, or whether higher level pragmatic reasoning permeates the process of semantic composition. This issue, in its turn, is of twofold importance for understanding the role of conversational implicature in linguistic communication.

1.2. Grice's approach and Marr's levels of analysis

What is the ontological status of Grice's approach to pragmatics? As noted before, several theorists have supposed that Grice's approach implies a certain view on utterance processing, according to which hearers actively calculate conversational implicatures. By contrast, Geurts and Rubio-Fernández (2015) argue that Grice's approach operates on an altogether different level than that of utterance processing. In motivating their proposal, Geurts and Rubio-Fernández appeal to Marr's (1982) distinction between computational and algorithmic levels of explanation.

The computational level of explanation describes what a system does and why, whereas the algorithmic level focuses on how that system achieves its goal. Put differently, the computational level describes a system from the outside, whereas the algorithmic level describes its internal processes. To illustrate, consider an explanation of the human heart. At the computational level, the human heart provides the body with oxygen and nutrients (what) in order to sustain life (why). At the algorithmic level, it achieves its goal by pumping blood around the body in a certain well-specified fashion (how).

Although the computational and the algorithmic levels constrain each other, there can be a significant amount of slack between them. Geurts and Rubio-Fernández illustrate this point using the example of syllogistic reasoning: Even though the set of valid types of syllogism is well defined at the computational level, its implementation in most humans is error prone and influenced by factors that are orthogonal to the notion of validity, for example, the content of the premises. Importantly, whether or not people judge a syllogism as valid is intuitively immaterial to the question whether or not it is in fact valid.

According to Geurts and Rubio-Fernández, Grice's approach operates at the computational level: It specifies the information hearers are entitled to infer (what) and how these inferences are justified (why). Notwithstanding a certain current in the literature, it does not say anything about the specifics of the inferential mechanism in our minds (how). Hence, although the literal-first hypothesis and its parallel processing competitor are mutually exclusive, both are compatible with Grice's approach to pragmatics.

This conclusion does not trivialize the question of how conversational implicatures are processed: Even though the computational and the algorithmic level do not designate each other, the connection between them is bound to be systematic. Determining the scope of Bott and Noveck's (2004) findings and pinpointing the locus of the processing cost associated with the derivation of scalar inferences will inform us about the nature of this systematic connection.

1.3. The origin of conversational implicature

Why did conversational implicatures emerge in the first place? Is it not more prudent and efficient to spell out exactly what we want to say rather than relying on the listener to work it out based on the literal interpretation of an utterance and the assumption that the speaker is cooperative? This question has received ample attention in the literature (e.g., Hauser, 1996; Horn, 2004; Pinker & Bloom, 1990; Pinker, Nowak, & Lee, 2008; Reboul, 2007).

According to Horn (2004), the evolution of communication is governed by a minimax principle that encourages interlocutors to minimize their efforts and to maximize their payoffs. For both interlocutors, payoffs are maximized if the speaker achieves her communicative goal. The speaker's effort may be equated to the length of the string she utters; the hearer's effort to the cognitive cost of the interpretative process (Horn, 1984; Martinet, 1962; Zipf, 1949). From the minimax perspective, then, conversational implicatures are not expected to emerge as long as the extra cognitive cost incurred by the hearer is greater than the effort saved by the speaker. Most authors have assumed that this is usually the case and hence tried to explain the presence of conversational implicatures by appealing to social constructs like face (Pinker et al., 2008) and manipulation (Reboul, 2007).

However, if it turns out that the extra cognitive cost for the hearer is restricted to scalar inferences, the emergence of conversational implicatures may be expected even within a communicative system that is governed by the minimax principle.

1.4. Outline

Why is the processing of scalar inferences so different from that of many other kinds of conversational implicature? To answer this question, we extended Bott and Noveck's (2004) paradigm to four kinds of inferences that are often explained as conversational implicatures: scalar inferences, free choice inferences, conditional perfection, and exhaustivity in "it"-clefts. We first discuss some of the previous work on the literal-first hypothesis. Based on this discussion, we consider two possible explanations for Bott and Noveck's results, describe the four kinds of inferences in more detail, and spell out how these inferences are predicted to behave. Afterward, we explain the details of our experiment and present the results.

2. Previous work

2.1. Metaphors

The publication of Grice's *Logic and conversation* was followed by a period of extensive work on the literal-first hypothesis. As a representative example, we discuss the work of Ortony et al. (1978) on metaphors. According to the "Gricean" view, the interpretation of metaphors proceeds from the observation that the literal meaning is defective (Grice, 1975; Searle, 1979). In order to test this view, Ortony and colleagues measured reading times for sentences such as:

(2) Regardless of the danger, the troops marched on.

In one condition, this sentence was preceded by a vignette about soldiers who found themselves in a dangerous situation. In the second condition, the preceding vignette was about children whose behavior angered their babysitter. So the first condition steered participants toward a literal interpretation of the target sentence; the second condition toward a metaphorical interpretation. In addition, Ortony and colleagues varied the length of the vignette: In one condition, it was just one sentence; in the other condition, it consisted of several sentences.

In the short vignette condition, all looked well for the literal-first hypothesis: Reading times were significantly longer in the condition that steered participants toward a metaphorical interpretation. However, when participants were provided with a more elaborate context, this difference disappeared and the metaphorical interpretation was accessed just as quickly as the literal interpretation.

Although the experimental details differ, analogous results were found for indirect questions (Clark, 1979; Gibbs, 1983), sarcasm (Gibbs, 1986a), metaphors (Gildea & Glucksberg, 1983; Glucksberg et al., 1982; Harris, 1976), and hyperbole (Deamer et al., 2010), all of which are often explained as conversational implicatures. Taken together, these results seem to falsify the literal-first hypothesis. However, there is one study whose results do support the literal-first hypothesis: Bott and Noveck (2004). We take a closer look at this study in the next section.

2.2. Scalar inferences

Following work by Noveck (2001) and Noveck and Posada (2003), Bott and Noveck (2004) investigated how long it takes for participants to determine the truth value of sentences like (3). Such sentences are true if "some" receives its literal interpretation as "at least some and possibly all," but they are false if a scalar inference is computed and "some" is interpreted with an upper bound.

(3) a. Some dogs are mammals.
    b. Some parrots are birds.

These scalar inferences are often explained as conversational implicatures (e.g., Gazdar, 1979; Horn, 1972 1984; Soames, 1982). According to such explanations, someone who hears (3a) may reason as follows: The speaker could have been more informative by using the alternative "All dogs are mammals." Why didn't she? Presumably because she does not believe that this alternative is true. Assuming, furthermore, that the speaker knows whether or not the alternative is true, it follows that the speaker takes the alternative to be false.

Participants in B&N's experiment read sentences like (3) in a forced reading experiment in which each word was flashed on the screen for 200 ms. They could provide their answer immediately after the final word was flashed on the screen. B&N measured the time between the onset of the final word and the moment one of the response buttons was pressed. Many participants were ambivalent about the truth of sentences like (3), varying their answer between structurally similar trials. Comparing the reaction times of these ambivalent participants, B&N found that it took them significantly longer to answer "false" than it took them to answer "true." This difference in reaction times was absent in a control condition in which the sentence was unambiguously true or false, as in (4). In this condition, "false" answers did not take significantly longer than "true" answers:

(4) a. Some birds are parrots.
    b. Some dogs are insects.

Taken together, these findings support the view that the enriched interpretation depends on the prior calculation of the literal interpretation. B&N's findings have since been replicated in a number of studies (e.g., Chemla & Bott, 2014; Chevallier et al., 2008; De Neys & Schaeken, 2007; Degen & Tanenhaus, 2011; Huang & Snedeker, 2009a, 2011; Tomlinson, Bailey, & Bott, 2013, but see Feeney, Scrafton, Duckworth, & Handley, 2004; Grodner, Klein, Carbary, & Tanenhaus, 2010, for apparently contradictory findings).

In sum, there is a marked difference in processing terms between scalar inferences, on the one hand, and indirect questions, sarcasm, idioms, metaphors, and hyperbole, on the other. What causes this difference? One potential explanation stems from the observation that scalar inferences are a variety of *quantity implicature*.

Quantity implicatures are so-called because their computation proceeds from the observation that the speaker could have been more informative, which is usually determined by comparing the informational value of an utterance with that of alternatives the speaker could have uttered. For example, in order to decide that (3a) is underinformative, it has to be compared with the alternative "All dogs are mammals." By contrast, the calculation of indirect questions, sarcasm, idioms, metaphors, and hyperbole starts with the recognition that the literal interpretation is either irrelevant (indirect questions) or nonsensical (sarcasm, idioms, metaphors, and hyperbole). In neither case is it necessary to reason about alternatives the speaker could have uttered.

Therefore, one might conjecture that the processing cost associated with the computation of scalar inferences is caused by having to reason about the speaker's beliefs based on premises that involve counterfactual situations. A suggestive observation in favor of this *counterfactual-reasoning hypothesis*, as we will call it, is that children in general have problems with tasks that involve counterfactual reasoning (Breheny, 2006; Proust, 2002; Riggs, Peterson, Robinson, & Mitchell, 1998; Wimmer & Perner, 1983). The counterfactual-reasoning hypothesis therefore predicts that children should also be less prone to compute scalar inferences. This prediction has been confirmed in a number of studies (e.g., Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Guasti et al., 2005; Huang & Snedeker, 2009b; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007, but see Barner, Brooks, & Bale, 2011; Katsos & Bishop, 2010).

It follows from the counterfactual-reasoning hypothesis that all varieties of quantity implicature should be associated with a processing cost, since the computation of these inferences by definition involves reasoning about what the speaker could have said. This prediction is challenged by findings reported in Chemla and Bott (2014), who extended B&N's paradigm to another kind of quantity implicature: *free choice inferences*. We discuss their study in the next section.

2.3. Free choice inferences

Free choice inferences are associated with utterances in which "or" is embedded under an existential or universal quantifier (e.g., Jennings, 1994; Kamp, 1973):

(5) Beulah may have tea or coffee.

Someone who utters (5) is likely to convey that Beulah is allowed to choose between having tea and having coffee. There has been some debate about the provenance of these free choice inferences. Some authors locate their source in the semantics of modal expressions (Barker, 2010; Merin, 1992) or the connective "or" (Geurts, 2005; Zimmermann, 2000). Other authors explain free choice inferences as a variety of quantity implicature (Fox, 2007; Franke, 2011; Geurts, 2010; Kratzer & Shimoyama, 2002; Schulz, 2005).

According to these implicature-based accounts, the derivation of free choice inferences proceeds along the following lines. A speaker who says (5) could have been more informative by using one of the following alternatives:

(6) a. Beulah may have tea.
    b. Beulah may have coffee.

Both of these alternatives are more informative than the literal meaning of (5), since a disjunction is true whenever one of its disjuncts is true but not vice versa.

Why didn't the speaker say, for example, (6a)? If she had used this alternative, she would have communicated not just that Beulah is allowed to have tea, but also that she is not allowed to have coffee. The reason the speaker avoided this alternative, then, is that she believes it is not the case that Beulah is allowed to have tea but not coffee. The same reasoning goes, mutatis mutandis, for (6b), which implies that the speaker believes that it is not the case that Beulah is allowed to have coffee but not tea. Together with the literal meaning of (5), this entails that Beulah is allowed to have tea and that she is allowed to have coffee.

There is an important difference between the computation of free choice inferences and scalar inferences. In the case of scalar inferences, it suffices for the hearer to reason about the literal meaning of the alternatives, whereas in order to compute free choice inferences, she also has to reason about what the speaker would have implied had she uttered one of the alternatives. So, assuming that the counterfactual-reasoning hypothesis is correct, that is, that reasoning about what the speaker could have said leads to a processing cost, and free choice inferences are predicted to lead to a similar or possibly even greater increase in reaction times.

Chemla and Bott (2014) provide substantive evidence against this prediction. For example, in their first experiment, participants were familiarized with a vignette according to which engineers were allowed to save artifacts and zoologists were allowed to save living beings. Afterwards, participants were presented with sentences such as (7), which are true on their literal interpretation, but false if free choice inferences are computed, since zoologists are not allowed to save artifacts and engineers are not allowed to save living beings:

(7) a. Beverly-the-zoologist is allowed to save a hammer or a lion.
b. Essie-the-engineer is allowed to save a kangaroo or a fork.

As in B&N's experiment, many participants were ambivalent about the truth of sentences like (7), varying their answer between structurally similar trials. C&B analyzed the response times of these ambivalent participants, but they did not find a significant difference between "true" and "false" answers. They also included a control condition involving sentences that were unambiguously true or false, such as (8). Comparing these two conditions, C&B found an interaction between condition (target or control) and response (true or false). In the control condition, "false" answers were significantly slower compared to "true" answers than in the target condition. So computing free choice inferences reduced the increase in response times that was present in the control condition.

(8) a. Federico-the-engineer is allowed to save a hammer or a fork.
b. Martina-the-engineer is allowed to save a lion or a kangaroo.

These findings appear to contradict the counterfactual-reasoning hypothesis, since the derivation of free choice inferences mirrors that of scalar inferences in that it proceeds from the observation that the speaker could have been more informative.

C&B tentatively suggest another explanation, which is intimately connected to the theory of alternatives propagated by Katzir (2007) (cf. also Fox & Katzir, 2010, but see Rooth, 1992, for a different approach). According to Katzir, alternatives are constructed from the uttered sentence by means of three elementary operations on its syntactic structure: deleting constituents, substituting constituents with elements from the lexicon, and replacing constituents with contextually given material. To illustrate, consider the following two pairs of sentences:

(9) a. Some dogs are mammals.
    b. All dogs are mammals.

(10) a. Federico-the-engineer is allowed to save a hammer or a lion.
     b. Federico-the-engineer is allowed to save a hammer.

Constructing (9b) as an alternative to (9a) requires substituting "some" with another constituent from the lexicon: "all." By contrast, the construction of (10b) from an utterance of (10a) involves deleting one of the disjuncts from the uttered sentence. In this respect, then, scalar inferences differ from free choice inferences.

Based on this observation, C&B suggest that the computation of a conversational implicature is only time consuming if constructing the required alternatives involves substituting elements from the uttered sentence with words from the lexicon. Since the alternatives needed for the derivation of free choice inferences can be constructed by removing constituents from the uttered sentence, the derivation of these inferences proceeds without a processing cost. In what follows, we refer to this explanation as the *lexical-access hypothesis*.

The lexical-access hypothesis receives support from a developmental study by Barner et al. (2011). In line with much previous evidence, these authors found that children are less likely to derive scalar inferences than adults. For example, almost all children indicated that the statement "Some animals are sleeping" was a correct description of a situation in which all animals were sleeping. However, Barner and colleagues also observed that almost all children considered the statement "Only some animals are sleeping" a correct description of the same situation. This finding suggests that children fail to see that "All animals are sleeping" is an alternative to the statement with "some." Assuming that the reason why children compute fewer scalar inferences than adults is the same reason why adults who compute scalar inferences take longer to respond, and it follows that B&N's results are caused by the need to construct alternatives by consulting the lexicon.

Alternatively, C&B's results may be construed as evidence against the view that free choice inferences are a variety of quantity implicature in the first place. Indeed, there are a number of other observations that cast doubt on this assumption (cf. Chemla, 2009; Geurts, 2010; Jennings, 1994; van Tiel, 2012; Tieu, Romoli, Zhou, & Crain, 2015). Under this assumption, the counterfactual-reasoning hypothesis can be salvaged in spite of C&B's results.

In order to arrive at a more decisive interpretation of the previous results, we extend the scope of inquiry to two kinds of inferences that have hitherto received less attention in the theoretical and experimental literature: conditional perfection and exhaustivity in "it"-clefts. Both of these inferences are often viewed as varieties of quantity implicature, that is, their derivation involves reasoning about what the speaker could have said. In neither case, however, are alternatives constructed by substituting constituents in the uttered sentence with elements from the lexicon, thus allowing a direct comparison of the counterfactual-reasoning and lexical-access hypotheses. In the next section, we consider these two inference types in some more detail.

3. Extending the scope

3.1. Conditional perfection

The notion of conditional perfection was popularized by Geis and Zwicky (1971) based inter alia on the following sentence:

(11) If you mow the lawn, I will give you five dollars.

On its literal interpretation, someone who utters this sentence implies that mowing the lawn is a sufficient condition for receiving five dollars. In other words, it is impossible that the hearer mows the lawn, but fails to receive five dollars. According to Geis and Zwicky, however, an utterance of this sentence normally implies, in addition, that mowing the lawn is a necessary condition for receiving five dollars. That is, the speaker will usually be taken to imply that it is impossible for the hearer to receive five dollars if she does not mow the lawn.

Several authors have argued that Geis and Zwicky's characterization is too strong (e.g., von Fintel, 2001, unpublished data; Franke, 2009; Geurts, 2010; Lilje, 1972). According to these authors, someone who utters (11) does not normally convey that mowing the lawn is the only course of action that would lead her to giving the hearer five dollars. For example, the speaker might also give five dollars if the hearer cleans up the garage or does the dishes. Instead, these authors assume that an utterance of (11) implies the weaker inference that the speaker will not give five dollars unconditionally. It is this weak variety of conditional perfection that we will test in our experiment.

There has been much debate about the provenance of conditional perfection. In his overview article, van der Auwera (1997) categorizes the various proposals that have been made. Of the 50 articles he reviews, 35 argue that conditional perfection is due to a quantity implicature. Eight articles go against this explanation, and the remaining seven articles are indeterminate. Much of the literature since van der Auwera's overview side with the implicature-based view (e.g., von Fintel, 2001; Franke, 2009; Geurts, 2010; Horn, 2000; Levinson, 2000). It therefore seems safe to claim that there is a broad consensus about the status of conditional perfection as a kind of quantity implicature.

How to compute the weak variety of conditional perfection as a quantity implicature? Suppose someone utters (11). She could have been more informative by saying "I will give you five dollars." Why didn't she? Presumably because she will not give the hearer five dollars unconditionally. According to this procedure, deriving conditional perfection involves reasoning about what the speaker could have said based on alternatives that are generated by deleting constituents in the speaker's utterance. In both respects, conditional perfection mirrors free choice inferences.

3.2. *Exhaustivity in "it"-clefts*

The following sentence illustrates exhaustivity in "it"-clefts:

(12) It was a pizza that Mary ate.

On its literal interpretation, someone who utters this sentence implies that Mary ate a pizza. This interpretation is compatible with a situation in which Mary ate other things besides a pizza. However, in most situations, someone who utters (12) will be taken to imply that Mary did not eat anything other than a pizza. It is the latter inference that is referred to as exhaustivity in "it"-clefts.

As for free choice inferences, there is no consensus about the provenance of exhaustivity in "it"-clefts. Although the implicature-based approach seems to be the most popular (e.g., Byram-Washburn, 2012; DeClerck, 1988; Dufter, 2009; Hartmann & Veenstra, 2013; Horn, 1981; Pavey, 2004; Vallduvì, 1993), alternative proposals locate the source of exhaustivity in the semantics of the cleft construction (Atlas & Levinson, 1981; Bolinger, 1972; Hedberg, 2000; Kiss, 1998), a conventional implicature (Halvorsen, 1978), or a presupposition (Büring & Križ, 2013; Percus, 1997). However, much of the recent experimental evidence speaks in favor of an implicature-based account (Destruel & Farmer, 2015; DeVeaugh-Geiss, Zimmermann, Onea, & Boell, 2015; Drenhaus, Zimmermann, & Vasishth, 2011; Heizmann, 2007; Onea & Beaver, 2011). As we will see later, our own results present further evidence in this respect.

One way of deriving exhaustivity in "it"-clefts as a quantity implicature is by reasoning as follows: If the speaker believed that Mary ate something other than a pizza, she would have mentioned it. Since she did not, it follows that she believes that Mary did not eat anything other than a pizza. According to this proposal, the derivation of exhaustivity in "it"-clefts involves reasoning about what the speaker could have said without recourse to alternatives altogether (Geurts, 2010).

In order to decide between the counterfactual-reasoning and lexical-access hypotheses, we compared the processing times of scalar inferences, free choice inferences, conditional perfection, and exhaustivity in "it"-clefts in an integrated experimental design. In the next section, we discuss the details of the experiment. Afterward, we outline the predictions of the two hypotheses.

4. This study

4.1. *The task*

Instead of B&N's and C&B's sentence verification tasks, we employed a sentence–picture verification task. This change in methodology has a practical motivation: For some of the inferences we investigated, it was difficult to find sentences that were introspectively true on their literal interpretation, but false if the inference in question was computed. The sentence–picture verification task avoids this difficulty. Moreover, it has already been employed in several studies on scalar inferences (e.g., Barner et al., 2011; Geurts & Pouscoulous, 2009; Geurts & van Tiel, 2013; Guasti et al., 2005; Papafragou & Musolino, 2003; Pouscoulous et al., 2007). Degen and Tanenhaus (2011, 2015a) replicated B&N's results using the sentence–picture verification task.

Each trial in our experiment started with a sentence that was followed by a picture. Participants had to indicate if the sentence was a good description of the picture that followed it. All of the situations consisted of a number of colored geometrical shapes. Four colors—red, green, blue, and yellow—and three kinds of geometrical shapes—squares, circles, and triangles—were used. The target sentences were followed by three types of situations: a target situation, in which the sentence was literally true, but false if the inference in question was derived, and two control situations, in which the sentence was unambiguously true or false. These control situations were included to compensate for possible response biases (Carpenter & Just, 1975; Clark & Chase, 1972). An overview of the target sentences and the corresponding situations is provided in Fig. 1. We will now discuss these conditions in more detail.

4.2. Scalar inferences

An example target sentence for scalar inferences is provided below, alongside its important characteristics. The logical form shows the assumed underlying representation of the target sentence used in the experiment. The literal interpretation paraphrases the truth conditions of the logical form. The alternative is constructed by substitution. The target inference paraphrases the information obtained by reasoning about the content of the alternative. Together with the literal interpretation, this produces the enriched interpretation. For clarification, linguistic forms are marked with quotation marks.

Target sentence:	"Some of the shapes are red."
Logical form:	$\exists x[\textsc{shape}(x) \wedge \textsc{red}(x)]$
Literal interpretation:	There is at least one red shape.
Alternative:	"All of the shapes are red."
Constructed by:	Substitution
Target inference:	Not all of the shapes are red.

The color term was varied between items. In the target situation, all of the shapes were red. In this situation, the sentence was true on its literal interpretation, but false if the target inference was derived. In the first control situation, some but not all of the shapes were red, in line with both the literal and the enriched interpretation of the target sentence. In the second control situation, none of the shapes were red, thus contradicting both possible interpretations.

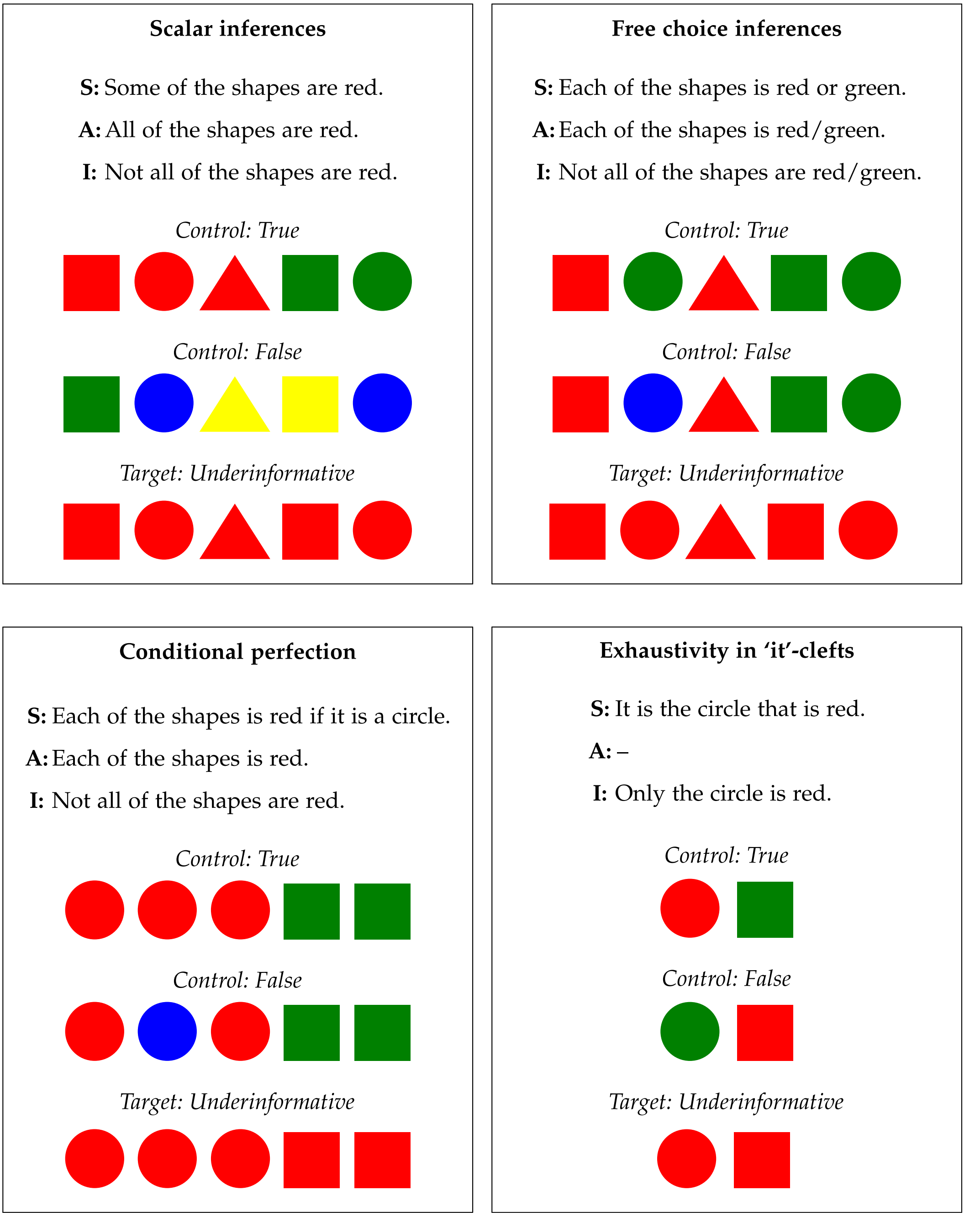


Fig. 1. Examples of target sentences for each of the four kinds of inferences. In the first picture, the sentence is unambiguously true; in the second picture, unambiguously false; and in the third picture, the truth value depends on whether the target inference is computed. **S**: Target sentence. **A**: Alternative. **I**: Target inference.

4.3. Free choice inferences

The characteristics of free choice inferences are described below:

Target sentence:	"Each of the shapes is red or green."
Logical form:	$\forall x[\textsc{shape}(x) \rightarrow [\textsc{red}(x) \vee \textsc{green}(x)]]$
Literal interpretation:	No shape is neither red nor green.
Alternative:	"Each of the shapes is red/green."
Constructed by:	Deletion
Target inference:	Not all of the shapes are red/green.

The color terms were varied between items. Note that the target sentence for free choice inferences differs from the one used by C&B. In our experimental sentence, "or" occurred embedded under a universal quantifier over individuals rather than under an existential quantifier over deontically accessible possible worlds. From a theoretical viewpoint, however, the computation of the free choice inferences proceeds along similar lines: The listener concludes from an utterance of the target sentence that the speaker considers the alternatives "Each of the shapes is red" and "Each of the shapes is green" to be false. Assuming that the speaker is truthful, it follows that there are both red and green shapes. Unlike in C&B's example, it suffices to reason about the literal meaning of the alternatives to compute free choice inferences in this example (Fox, 2007; Geurts, 2010; Klinedinst, 2007, but see Crnič, Chemla, & Fox, 2015, for an argument that these cases also involve reasoning about the pragmatic meaning of the alternatives).

There are two motivations for our departure from C&B. First, it was difficult to straightforwardly encode permission in pictures. Second, testing a different kind of target sentence provides a means to gauge the robustness and generality of C&B's findings for free choice inferences on the whole. See Data S1 in the Supplementary Material for results of an experiment that stays closer to C&B's design.

In the target situation, either each of the shapes was red or each of the shapes was green. In these situations, the target sentence was true on its literal interpretation, but false if free choice inferences were computed to arrive at the conclusion that there were both red and green circles. The first control situation showed a mixture of red and green circles, thus verifying both the literal and the enriched interpretation. The second control situation contained one shape that was neither red nor green, which falsified the target sentence on both of its interpretations.

4.4. Conditional perfection

The characteristics of conditional perfection are outlined below:

Target sentence:	"Each of the shapes is red if it is a circle."
Logical form:	$\forall x[\textsc{circle}(x) \rightarrow \textsc{red}(x)]$
Literal interpretation:	No circle is not red.
Alternative:	"Each of the shapes is red."
Constructed by:	Deletion
Target inference:	Not all of the shapes are red.

The color and shape terms were varied between items. All situations contained three circles and two squares. In the target situation, all of the shapes were colored red, so that the sentence was true on its literal interpretation but false if the conditional was perfected. In the first control situation, each of the circles was red while each of the squares was a different color, in line with both the literal and the enriched interpretation. In the second control situation, at least one of the circles was not red, thus falsifying the sentence.

4.5. Exhaustivity in "it"-clefts

The characteristics of exhaustivity in "it"-clefts, the fourth and final kind of inference included in the experiment, are laid out below. Recall that the computation of exhaustivity in "it"-clefts proceeds without recourse to alternatives.

Target sentence:	"It is the circle that is red."
Logical form:	$\exists! x[\textsc{circle}(x) \wedge \textsc{red}(x)]$
Literal interpretation:	The circle is red.
Alternative:	–
Constructed by:	–
Target inference:	Only the circle is red.

The color and shape terms were varied between items. All situations contained one circle and one square. In the target situation, both shapes were red. In this situation, the exhaustivity inference was false. In the first control condition, the circle was red and the square a different color. In the second control condition, the square was red and the circle a different color. These situations unambiguously verify and falsify the sentence.

4.6. Predictions

In the previous sections, we have discussed two possible explanations for B&N's finding that the derivation of scalar inferences, unlike that of several other varieties of conversational implicature, is associated with a processing cost.

According to the counterfactual-reasoning hypothesis, the processing cost is due to the difficulty of recognizing that the literal interpretation is underinformative, since doing so involves reasoning about what the speaker could have said. In order to accommodate C&B's finding that the derivation of free choice inferences is not associated with a processing cost, this explanation has to assume that these inferences are not a variety of quantity implicature.

If this hypothesis is correct, we expect to find that conditional perfection and exhaustivity in "it"-clefts, assuming that the latter is indeed due to a quantity implicature, pattern with scalar inferences in leading to a delay in response times. Concretely, this delay in response times entails the presence of an interaction between condition (control or target) and response (true or false) for these two types of inferences, such that "false" responses take significantly longer than "true" responses in the target condition when compared to the control condition.

The lexical-access hypothesis supposes that whether or not the computation of a conversational implicature is associated with a processing cost depends on the structural characteristics of the required alternatives. Specifically, a processing cost is expected if the alternatives are constructed by substituting constituents in the uttered sentence with words from the lexicon. This explanation straightforwardly extends to C&B's results: Free choice inferences are not associated with a processing cost since the alternatives needed for their derivation can be constructed by removing constituents from the uttered sentence.

If the lexical-access hypothesis is correct, the computation of conditional perfection and exhaustivity in "it"-clefts should not lead to an increase in response times. Indeed, both kinds of inferences should pattern with free choice inferences and differ significantly from scalar inferences. Generalizing C&B's results for free choice inferences, this hypothesis predicts the presence of an interaction between condition (control or target) and response (true or false) for these two types of inferences, such that "true" responses take significantly longer than "false" responses in the target condition when compared to the control condition.

In Fig. 2, the predictions made by the counterfactual-reasoning and lexical-access hypotheses are displayed, assuming that we replicate B&N's and C&B's findings for scalar inferences and free choice inferences. To explain, consider the case of conditional perfection according to the counterfactual-reasoning hypothesis. The line indicates that the difference between the response times in target and control conditions is smaller for "true" answers than for "false" answers. More specifically, in the case of "false" answers, responses times in the target condition are significantly slower than in the control condition when compared to the difference between these conditions in the case of "true" responses. In more intuitive terms, the line indicates that the counterfactual-reasoning hypothesis predicts a processing cost for the derivation of conditional perfection.

In this discussion, we have passed over the question of whether the increase in response times is present in the reading times, decision times, or both. We do not have specific expectations either way, but it will be interesting to see at what point participants engage in pragmatic reasoning.

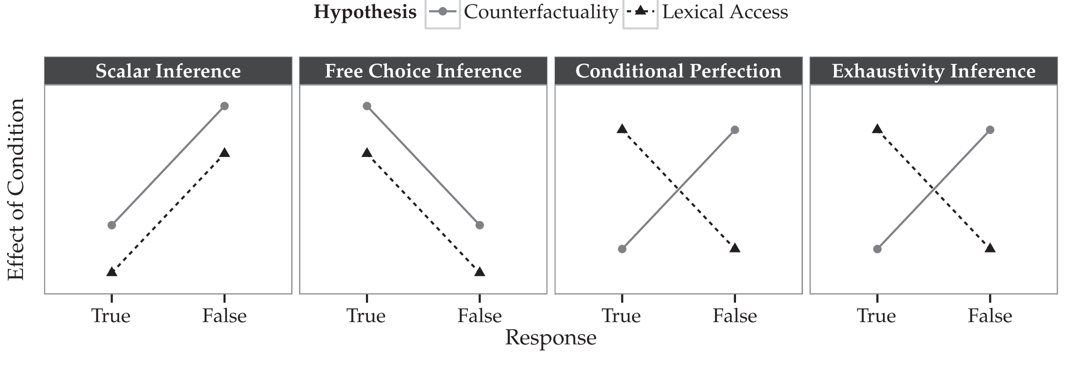


Fig. 2. Interaction effects (target minus control) predicted by the counterfactual reasoning hypothesis (black) and the lexical access hypothesis (gray).

In the next section, we outline the details of the sentence–picture verification task conducted to test these predictions. Afterward, we discuss the results.

5. The experiment

5.1. Participants

The experiment was hosted on the Ibex Farm. Forty-five participants were drafted on Amazon's Mechanical Turk ($M_{\text{age}} = 30$; range = 19–59; 24 females).[1] Only workers with an IP address from the United States were eligible for participation. In addition, these workers were asked to indicate their native language. Payment was not contingent on their response to this question. All participants were native speakers of English. One participant indicated that both English and Spanish were her native languages. We included this participant in this analysis we discuss below.

5.2. Materials

The experiment consisted of 60 trials in total and included four types of sentences corresponding to the four types of inferences:

(13)	a. Some of the shapes are C.	*Scalar inference*
	b. Each of the shapes is C_1 or C_2.	*Free choice inference*
	c. Each of the shapes is C if it is a S.	*Conditional perfection*
	d. It is the S that is C.	*Exhaustivity in "it"-clefts*

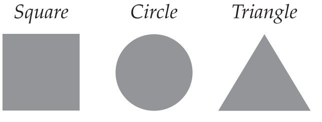
C was varied between "red," "green," "blue," and "yellow." S was varied between "square," "circle," and "triangle." The pictures for the first three types of inferences always consisted of five shapes. The pictures for exhaustivity in "it"-clefts always consisted of two shapes.

For each inference type, three kinds of situations were constructed: two control situations and one target situation. In the first control situation, the sentence was unambiguously true; in the second control situation, it was unambiguously false; in the target situation, its truth value depended on whether a conversational implicature was derived. See Fig. 1 for examples of these situations. Each kind of situation occurred five times for each inference type. The order of the items was randomized for each participant.

5.3. Procedure

On each trial, the target sentence was displayed first. Participants were instructed to press the space bar as soon as they had read and understood the sentence. Thereupon, the sentence disappeared and was replaced by a picture. Participants had to decide as quickly as possible whether the sentence was true or false as a description of the depicted situation, and had to register their decision by pressing one of two keys. Thereupon, the picture disappeared and was replaced by the message "(Press the space bar to continue.)." Upon pressing the space bar, the next trial commenced. The full instructions went as follows:

In this experiment you are going to read sentences. Each sentence is followed by a picture. It's your task to decide if the sentence is a good description of the picture that follows it. Each picture shows a number of colored shapes. There are three possible shapes:



First the sentence will appear. Press the space bar once you have read it. After that the sentence disappears and is replaced by the picture. Press '1' if you consider the sentence a good description of the picture; otherwise press '0'.

We are interested in your spontaneous responses, so don't think too long before answering.

Reading times were recorded from sentence onset to the point at which the space bar was pressed. Decision times were recorded from situation onset to the point at which the "1" or "0" key was pressed.[2]

5.4. Data treatment

Six participants were removed for making mistakes on more than 10% of the control items. This cutoff criterion was chosen because there was a significant gap between the participants that were removed (more than 9 errors) and the participants that were retained (between 0 and 4 errors). The average rate of mistakes for the participants that were retained was 3.5%. These mistakes were removed from analyses of reading and decision times but are plotted in Fig. 3. In addition, trials with a reading or decision time faster than 200 milliseconds or slower than three times the standard deviation from the mean by participant and condition were removed from the analysis, which resulted in the removal of 2.2% of the data. We assume that these data points indicate that participants either accidentally pressed one of the response keys or did not pay full attention to the task at hand. The reading and decision times were logarithmized to reduce the positive skewness of their distribution.

5.5. Inferential analysis

The design contained three independent variables: inference type (scalar inference, free choice inference, conditional perfection, or exhaustivity in "it"-clefts), condition (control or target), and response (true or false).

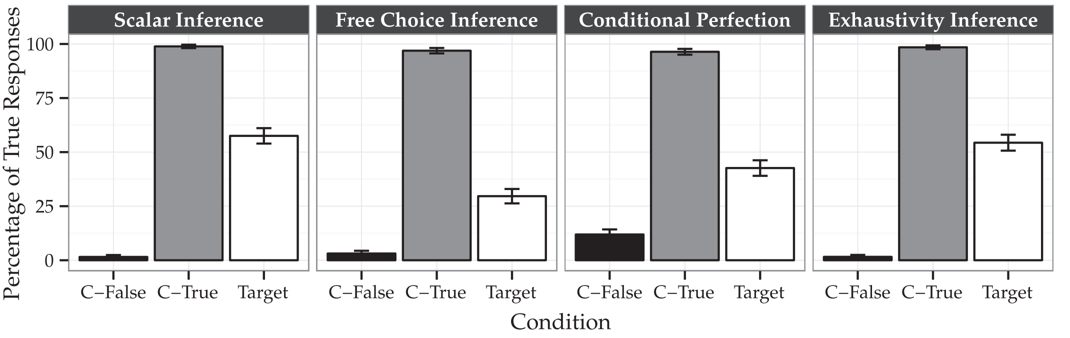


Fig. 3. Percentage of “true” responses for each inference type in control (C-True and C-False) and target situations.

The results were analyzed by fitting linear or binomial mixed models predicting reading times, decision times, and choice proportions. These models were constructed in R, a programming language and environment for statistical computing (R Development Core Team, 2006) using the `lme4` package (Bates & Maechler, 2009). The corresponding *p*-values were calculated by approximating denominator degrees of freedom using the Satterthwaite procedure as implemented in the `lmertest` package (Kuznetsova, Bruun Brockhoff, & Haubo Bojensen Christensen, 2013).

In all analyses, random slopes and intercepts were included for participants and items (Barr, Levy, Scheepers, & Tily, 2013). In some analyses, the random effects structure had to be simplified because the maximal model failed to converge. Multiple comparisons were carried out using the `multcomp` package (Hothorn, Bretz, & Westfall, 2008). The analyses of reading times and decision times included trial number as a fixed factor.

5.6. Choice proportions

The percentages of “true” responses in each condition are summarized in Fig. 3. For scalar inferences (58% literal responses), conditional perfection (43%), and exhaustivity in “it”-clefts (54%), literal responses were roughly as frequent as pragmatic responses. Participants in the case of free choice inferences had a pronounced preference for pragmatic responses: 70% of the responses were pragmatic. A binomial mixed model was constructed to predict responses (true or false) based on inference type (scalar inference, free choice inference, conditional perfection, or exhaustivity in “it”-clefts). Post hoc comparisons indicated that all four choice proportions differed significantly from each other (p’s $< .02$) except for the choice proportions for scalar inferences and exhaustivity in “it”-clefts ($\beta = 0.33$, $SE = 0.28$, $Z = 1.18$, $p = .636$). There was no effect of trial number on responses to target items ($\beta = 0.00$, $SE = 0.01$, $Z < 1$). That is, the number of pragmatic responses did not increase or decrease throughout the experiment.

Some participants consistently gave literal or pragmatic responses. For scalar inferences (18 consistently literal responders and 11 consistently pragmatic responders) and exhaustivity in "it"-clefts (12/11), the consistently literal responders outnumbered the consistently pragmatic responders. For conditional perfection (7/13) and free choice inferences (6/22), the opposite pattern was observed. Fig. 4 shows that the proportions of literal responses were distributed bimodally for each inference type, although the bimodality was slightly more pronounced for scalar inferences and free choice inferences than for exhaustivity in "it"-clefts and, in particular, conditional perfection. Overall, there were seven participants who provided consistent answers across all inference types. Three gave exclusively literal responses; four exclusively pragmatic responses.

One might wonder whether participants were consistent across different inference types, that is, whether participants gave a comparable number of pragmatic responses for each inference type. To test this, we calculated the product–moment correlation between the number of pragmatic responses for each pair of inference types. The results of this analysis are summarized in Fig. 5.

The correlation was significant across the board. The number of pragmatic responses for scalar inferences correlated significantly with the number of pragmatic responses for free choice inferences ($r = .46$, $p = .003$), conditional perfection ($r = .46$, $p = .003$), and exhaustivity in "it"-clefts ($r = .73$, $p < .001$). The number of pragmatic responses for free choice inferences correlated significantly with the number of pragmatic responses for conditional perfection ($r = .44$, $p = .005$) and exhaustivity in "it"-clefts ($r = .44$, $p = .005$). Lastly, the number of pragmatic responses for conditional perfection correlated significantly with the number of pragmatic responses for exhaustivity in "it"-clefts ($r = .67$, $p < .001$).

Both the observation that participants are ambivalent about the truth value of the target sentence and the finding that the number of pragmatic responses correlates significantly with the number of pragmatic responses to scalar inferences and conditional perfection are in line with the assumption that free choice inferences and exhaustivity in "it"-clefts are varieties of quantity implicature.

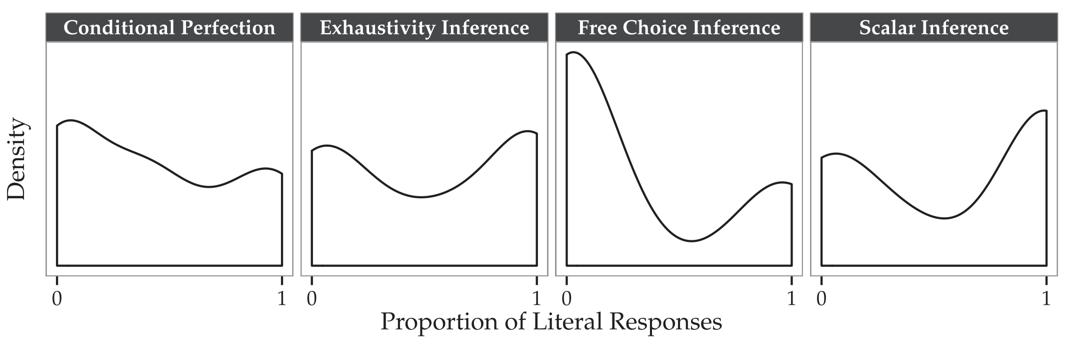


Fig. 4. Distribution of the proportions of literal responses for each of the four inference types.

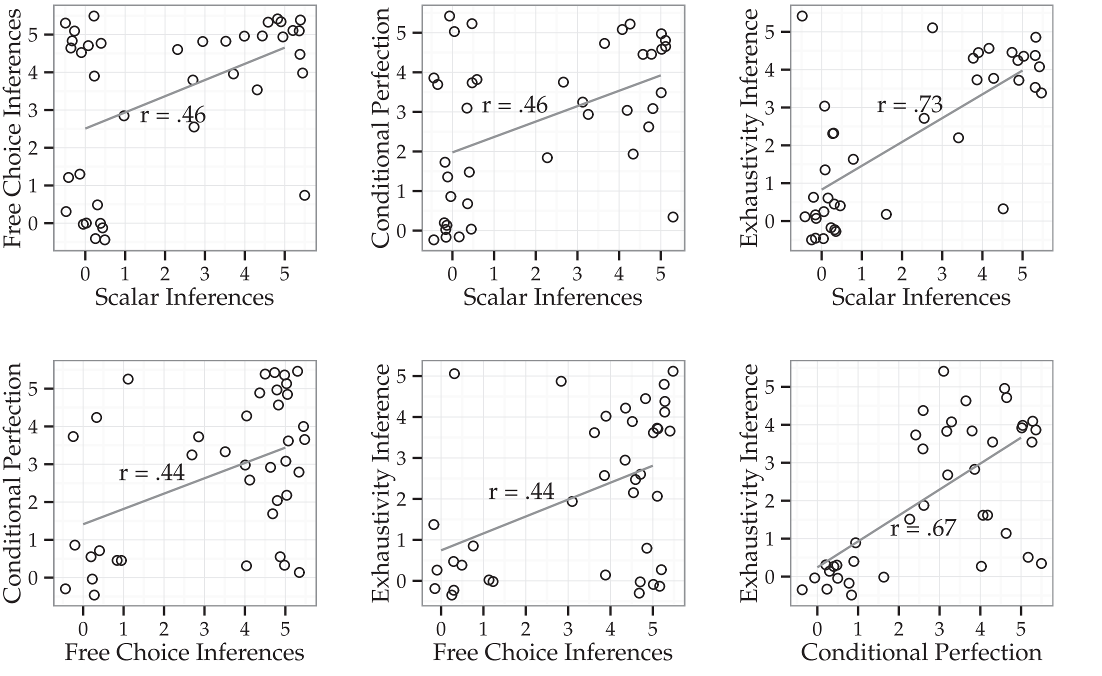


Fig. 5. The number of pragmatic responses for each pair of inference types including regression lines with product–moment correlation coefficients. Jitter added for purposes of illustration.

5.7. Reading times

Trial number had a significant effect in all of the analyses (all p's < .001): Reading times became significantly shorter throughout the experiment. Reading times were significantly different for each inference type (all p's < .002). There was no main effect of condition (control or target) or response (true or false) for any of the four types of inferences (all p's > .19). Importantly, the interaction between condition and response was also not significant for any of the four types of inferences (all t's < 1).

5.8. Decision times

The average decision times for each kind of inference are summarized in Fig. 6 and Table 1. Trial number had a significant effect in all of the subsequent analyses (all p's < .001): Decision times became significantly shorter throughout the experiment. Decision times for conditional perfection were significantly slower than for the other three inference types (all p's < .02). There were no significant differences in decision times for the other inference types (all p's < .19).

The interaction between condition (control or target) and response (true or false) was significant for each inference type: scalar inferences ($\beta = -0.29$, $SE = 0.12$, $t = -2.50$, $p = .027$), free choice inferences ($\beta = 0.27$, $SE = 0.07$, $t = 3.84$, $p < .001$), conditional

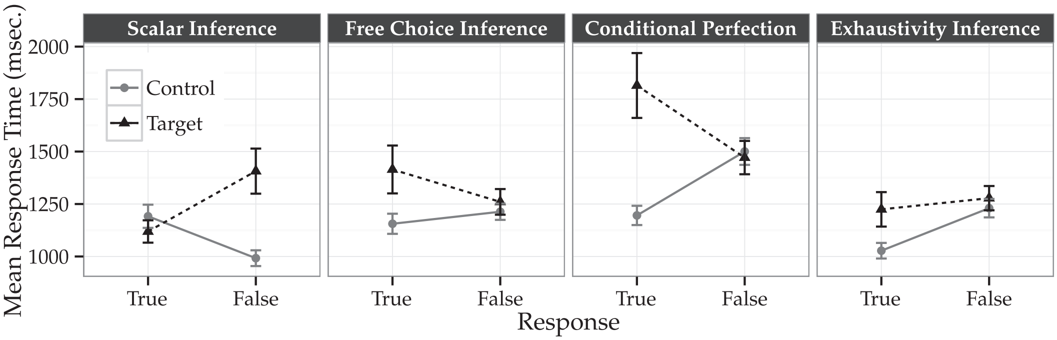


Fig. 6. Mean decision times for the four kinds of inferences. Error bars represent standard errors.

perfection (β = 0.33, *SE* = 0.08, t = 4.17, p < .001), and exhaustivity in 'it'-clefts (β = 0.18, *SE* = 0.07, t = 2.43, p = .019).

Observe that the direction of the interaction was different for scalar inferences compared to the other three types of inferences. In the case of scalar inferences, "false" responses were slower than "true" responses in the target condition when compared to the difference between these responses in the control condition. By contrast, in the cases of the other three types of inferences, "false" responses were faster than "true" responses in the target condition when compared to the difference between these responses in the control condition. In other words, in the cases of free choice inferences, conditional perfection, and exhaustivity in "it"-clefts, computing the inference facilitated decision times.

In order to determine if the interaction between condition and response was significantly different for different kinds of inferences, we analyzed the three-way interactions between condition, response, and kind of inference for each pair of inferences. The three-way interaction was significant for scalar inferences and each of the other three inferences: free choice inferences (β = −0.56, *SE* = 0.12, t = −4.54, p < .001), conditional perfection (β = −0.66, *SE* = 0.12, t = −5.52, p < .001), and exhaustivity in "it"-clefts (β = −0.46, *SE* = 0.11, t = −4.08, p < .001). In addition, there was a significant three-way interaction between conditional perfection and exhaustivity in "it"-clefts (β = −0.24, *SE* = −0.10, t = −2.40, p = .020). The remaining three-way interactions were not significant (both p's > .36).

One of our reviewers was concerned that some of the three-way interactions between condition, response, and inference type might be driven by our specific choice of control items. For that reason, we analyzed the interaction between response and inference type for the target condition alone. The results of these analyses should be interpreted with caution because they fail to take into account possible response biases.

The interaction between response and inference type was significant for scalar inferences and free choice inferences (β = −0.32, *SE* = 0.09, t = −3.52, p = .002), scalar inferences and conditional perfection (β = −0.25, *SE* = 0.10, t = −2.47, p = .015), and marginally significant for scalar inferences and exhaustivity in "it"-clefts (β = −0.15, *SE* = 0.08, t = −1.87, p = .069). In addition, it was significant for free choice inferences

Table 1
Means and standard errors for the decision times for the four kinds of inferences

	Target		Control	
	True	False	True	False
Scalar inference	1,119 (53)	1,407 (107)	1,192 (55)	992 (38)
Free choice inference	1,415 (114)	1,260 (61)	1,156 (48)	1,213 (39)
Conditional perfection	1,815 (155)	1,471 (79)	1,196 (46)	1,500 (64)
Exhaustivity inference	1,225 (82)	1,278 (58)	1,028 (37)	1,230 (44)

and exhaustivity in "it"-clefts ($\beta = 0.23$, $SE = 0.09$, $t = 2.49$, $p = .016$) and conditional perfection and exhaustivity in "it"-clefts ($\beta = 0.18$, $SE = 0.08$, $t = 2.22$, $p = .028$). The remaining interaction between free choice inferences and conditional perfection was not significant ($\beta = 0.09$, $SE = 0.10$, $t < 1$).

Focusing on the target condition, the difference between "true" and "false" responses was marginally significant for scalar inferences ($\beta = -0.18$, $SE = 0.09$, $t = -2.04$, $p = .058$) and significant for free choice inferences ($\beta = -0.18$, $SE = 0.09$, $t = -1.99$, $p = .049$). The difference was not significant for conditional perfection and exhaustivity in "it"-clefts (both t's < 1). Focusing on the control condition, the difference between "true" and "false" responses was not significant for scalar inferences ($\beta = 0.16$, $SE = 0.09$, $t = 1.79$, $p = .123$), it was marginally significant for free choice inferences ($\beta = -0.09$, $SE = 0.04$, $t = -2.27$, $p = .054$), and it was significant for the other two inference types (both p's < .01).

The results for the target condition were confirmed in a between-participants analysis. For this analysis, all participants were categorized as literal or pragmatic responder depending on their preferred response for a particular inference type. For example, someone who provided three pragmatic responses and two literal responses for scalar inferences was classified as a pragmatic responder for this inference type, and the results for her nonpreferred response were discarded from the analysis. A generalized linear model was constructed predicting logarithmized decision times in target conditions based on type of responder. The difference between literal and pragmatic responders was significant for scalar inferences ($\beta = -0.21$, $SE = 0.10$, $t = -2.17$, $p = .036$), but not for any of the other inference types (all p's > .28).

As discussed in Section 2.1, Ortony and colleagues found that the interpretation of a metaphorical expression was only associated with an increase in reading times if it was not sufficiently contextualized. One of our reviewers speculated that our data might bear evidence of a similar pattern. Although none of the sentences were properly contextualized, precedent might facilitate the processing of the pragmatic inferences. In order to evaluate this hypothesis, we repeated the within-participants and between-participants analyses of the logarithmized decision times in the target condition of each inference type, this time including the interaction between trial number and response as an independent factor. If precedent facilitates the processing of pragmatic inferences, we expect that decision times for "false" responses decrease with trial number at a faster rate than decision times for "true" responses, since the former involves pragmatic inferencing.

In the within-participants analysis, the interaction between trial number and response was not significant for any of the four inference types (all p's > .21). In the between-participants analysis, the interaction was marginally significant for conditional perfection ($\beta = -0.01$, $SE = 0.00$, $t = -1.76$, $p = .08$) in the expected direction; that is, the difference in decision times between "true" and "false" responses became smaller throughout the experiment. The interaction was not significant for any of the other inference types (all p's > .22). Hence, there is little evidence that precedent might have facilitated the processing of pragmatic inferences.

These findings can be summarized as follows: The computation of scalar inferences was associated with an increase in decision times. The results for the other three types of inferences differed significantly from the pattern for scalar inferences in that their computation facilitated decision times. In addition, there was a significant difference in the processing profiles of conditional perfection and exhaustivity in "it"-clefts, with the former having a greater facilitation effect than the latter.

Before discussing these results, we briefly want to consider a factor that might have had a confounding effect on the results: the complexity of the verification procedure. It has been suggested that the reason pragmatic responses take longer than literal responses for scalar inferences might be that it is more difficult to evaluate the enriched interpretation than the literal interpretation (Bott, Bailey, & Grodner, 2012; Geurts, 2010). To illustrate, consider the target picture for scalar inferences in Fig. 1. In order to establish that the literal meaning is correct, it suffices to inspect the first two circles. Since these are red, it follows that the sentence is true on its literal interpretation. In order to determine the truth value of the enriched interpretation, however, it is necessary to inspect each of the five circles. Perhaps, that is why pragmatic responses took longer than literal responses.

Note that there is a priori evidence against this hypothesis. For conditional perfection and exhaustivity in "it"-clefts, the enriched meaning of the target sentence was also more difficult to evaluate than the literal meaning. However, the computation of these two types of inferences did not lead to an increase in decision times and even facilitated the decision process. For a more comprehensive evaluation of the complexity hypothesis, we determined for each condition the minimum number of shapes that had to be evaluated in order to establish the truth value of the sentence. Visual complexity thus operationalized was not a significant predictor of decision times ($\beta = -0.01$, $SE = 0.01$, $t < 1$).

6. General discussion

6.1. Summary

In this study, we investigated the processing of four kinds of inferences that are often explained as quantity implicatures: scalar inferences, free choice inferences, conditional perfection, and exhaustivity in "it"-clefts. Replicating Bott and Noveck (2004), we found that the computation of scalar inferences caused an increase in decision times. Replicating Chemla and Bott (2014), the computation of free choice inferences was not associated with such an increase. Indeed, the within-participants analysis suggested that the computation of free choice inferences even facilitated the decision process. Conditional perfection and exhaustivity in “it”-clefts patterned with free choice inferences in this respect.

We discussed two possible explanations for B&N’s finding that the computation of scalar inferences leads to an increase in response times, given previous findings suggesting that other varieties of conversational implicature, such as indirect questions, sarcasm, idioms, metaphors, and hyperbole are computed without a processing effort. According to the counterfactual-reasoning hypothesis, the processing cost for scalar inferences is induced by having to construct and reason with alternatives the speaker could have uttered. The lexical-access hypothesis supposes that this processing cost depends on the structural characteristics of the alternatives. Specifically, a processing cost occurs only if the construction of alternatives requires substituting constituents in the uttered sentence with words from the lexicon. The predictions of the two hypotheses and the outcome of the experiment are summarized in Fig. 7.

The results of this experiment are in line with the lexical-access hypothesis. Conditional perfection and exhaustivity in “it”-clefts require reasoning about what the speaker could have said. The counterfactual-reasoning hypothesis therefore predicts that both types of inferences should be associated with a processing cost. This prediction is falsi-

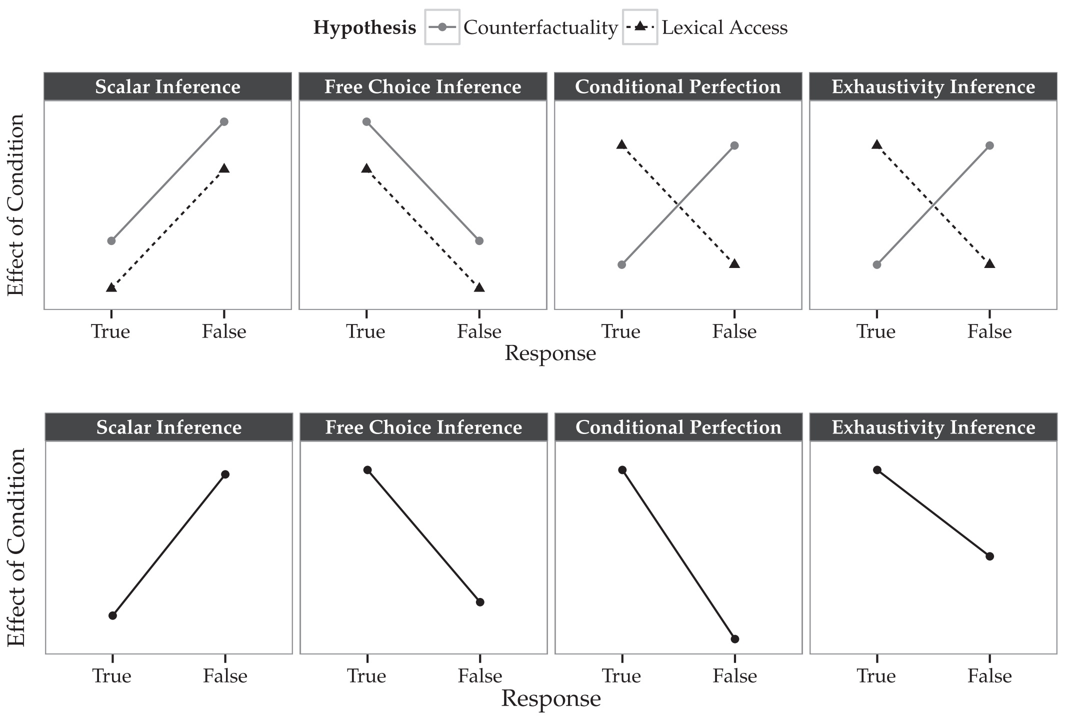


Fig. 7. Interaction effects (target minus control) predicted by the counterfactual-reasoning hypothesis and the lexical access hypothesis above the interaction effects estimated by the mixed models reported in section 5.8.

fied by the results of our experiment. By contrast, for neither of these inferences, is it necessary to construct alternatives by substituting constituents in the uttered sentence with words from the lexicon. The lexical-access hypothesis thus correctly predicts the absence of a processing cost for these two inference types.

One intriguing wrinkle in the data is that the size of the difference between "true" and "false" responses was greater for conditional perfection than for exhaustivity in "it"-clefts. This suggests that there may be even more subtle differences in processing costs depending on the type of alternatives. Recall that the computation of exhaustivity in "it"-clefts proceeds without recourse to alternatives, whereas the derivation of conditional perfection involves constructing alternatives by deleting constituents in the uttered sentence. Perhaps, this difference underlies the different processing profiles of these two inferences. A problem with this explanation is that there was no difference in the processing profiles of exhaustivity in "it"-clefts and free choice inferences, which also involve alternative construction through deletion. One potential explanation is that sentences with embedded occurrences of "or" evoke multiple alternatives corresponding to the individual disjuncts without "or," whereas sentences licensing conditional perfection evoke only one alternative.

Caution is warranted when evaluating such post hoc explanations, however, since these were not tested directly in our experiment. Moreover, constructing an experiment that teases apart these subtle differences will provide a difficult challenge. What is clear, however, is that there seem to be additional factors at work in determining the processing cost of a particular inference. See Data S1 in the Supplementary Material for some more discussion of this point.

It follows from the lexical-access explanation that alternatives are structurally defined in terms of elementary operations on the uttered sentence, in particular, deletion and replacement (cf. Fox & Katzir, 2010; Katzir, 2007, pace Rooth, 1992) and that sentence processing is sensitive to the details of these structural modifications.

The status of free choice inferences and exhaustivity in "it"-clefts as a variety of quantity implicature has been disputed in the literature. Our results provide tentative evidence that the implicature-based approach is on the right track for both types of inferences. In both cases, participants were hesitant about whether to judge the target sentence true or false in a situation in which the inference was false. In this respect, both types of inferences behave just like scalar inferences and conditional perfection, whose status as a variety of quantity implicature is less controversial. Perhaps more convincingly, participants who responded pragmatically to scalar inferences and conditional perfection also tended to respond pragmatically to free choice inferences and exhaustivity in "it"-clefts. This suggests that responses were driven by the same mechanism for each of the four inference types.

Of course, these observations should not be construed as incontrovertible evidence for the implicature-based approach. First, it may be the case that further theoretical work provides a more compelling and inclusive explanation of free choice inferences and exhaustivity in "it"-clefts. Second, it is altogether unclear what the precise processing predictions for alternative approaches are. Third, the aforementioned correlations could be driven by response biases. Defusing this counterargument would involve testing items with different kinds of ambiguities and see if responses to these items correlate with those for the target items, as predicted if the results are driven by response biases. Taken together, these considerations limit the scope of our conclusion to claiming that the data we gathered for free choice inferences and exhaustivity in “it”-clefts are in line with what would be predicted on an implicature-based approach.

6.2. Grice’s approach and Marr’s levels of analysis

The results of our experiment provide further evidence against construing Grice’s approach as an account of utterance processing. According to Grice, conversational implicatures are calculated using an argument based on the literal interpretation of an utterance and the assumption that the speaker is cooperative. Mapping this definition to utterance processing would entail that the literal interpretation of an utterance precedes the enriched interpretation. This prediction is falsified by the processing data for free choice inferences, conditional perfection, and exhaustivity in “it”-clefts. For each of these inferences, the enriched interpretation was accessed at least as quickly as the literal interpretation.

As discussed before, however, it would be an exegetical and conceptual mistake to interpret Grice’s approach as an account of utterance processing. Grice’s approach is situated at the computational level, explaining what inferences listeners are licensed to draw and why. It should not be confused for an algorithmic account of how these inferences are computed. Our results indicate that such an algorithmic account should allow for parallel processing of enriched meanings (e.g., Chierchia, 2004; Geurts, 2010; Levinson, 2000; Récanati, 1995).

There is thus a substantial difference between the serial computation procedure postulated at the computational level and the parallel processing that occurs at the algorithmic level. How problematic is this difference for Grice’s approach?

One method of bridging the gap between Grice’s approach and the processing data is to assume that conversational implicatures can become conventionalized (Horn & Bayer, 1984; Levinson, 2000; Morgan, 1978; Searle, 1975). This assumption entails that certain conversational implicatures are read off directly from the form of the uttered sentences instead of being actively computed, in a similar sense as the meaning of a written string may be read off directly from its form rather than being determined by converting letters to phonemes (e.g., Coltheart, 2003). For example, upon encountering a sentence of the form “P if Q,” the hearer can immediately infer that the speaker does not believe P to be necessarily true.

For this account to work, it remains to be explained why the computation of scalar inferences is still associated with a processing cost. Two such explanations suggest themselves. First, it may be the case that sentences of the form “Some A are B” evoke an underspecified inference “Not Q A are B” which needs to be parametrized by inserting a suitable replacement for “Q.” This explanation reflects the observation that a sentence with “some” can license different inferences depending on the context. For example, “Some A are B” may imply that not all A are B, that not most A are B, that not many A are B, or even that it is not just one, two, or three A that are B (Degen & Tanenhaus, 2011; Franke, 2014; van Tiel, 2014a; van Tiel & Geurts, 2014). This proposal complements the lexical-access hypothesis: Scalar inferences are associated with a processing cost because their computation involves parametrization by drawing upon the lexicon. If so, it is predicted that the processing cost disappears in situations in which it is clear what the relevant parameters are.

An alternative explanation is based on the hypothesis that the processing cost of a conventionalized conversational implicature depends on the frequency of the enriched interpretation in natural conversations. The observation that the proportion of literal and pragmatic responses varied across inference types provides some support for the role of statistical regularities. Why were participants more likely to compute conditional perfection than scalar inferences, and scalar inferences than free choice inferences? Prima facie it seems difficult to find a structural basis for this difference, and it seems probable that various cues conspire in influencing rates of pragmatic inferencing, including statistical regularities (cf. Seidenberg, 1997).

According to this proposal, scalar inferences are associated with a processing cost because they are in general less frequent than the other three types of inferences. This explanation receives some support from work by Degen (2015), who extracted sentences with "some" from a corpus of natural language and asked participants to indicate how likely it is that they would derive the scalar inference. Her results suggest that scalar inferences are relatively infrequent in natural language. However, further work is needed to decide between these proposals and spell out the role of statistical regularities in pragmatic inferencing.

6.3. The origin of implicit meaning

What is the function of conversational implicature in linguistic communication? Why use the underspecified "some" rather than its unambiguous alternatives "at least some" and "some but not all"? It seems plausible that the speaker desires to keep her utterances as short as possible, especially if the distinction between the literal and the enriched interpretation is not particularly relevant for the purpose of the conversation. However, if the use of an underinfomative sentence results in a protracted reasoning process on the side of the hearer (B&N observed a difference of about 600 ms between literal and pragmatic responses for scalar inferences), the distribution of effort seems to inordinately favor the speaker.

Our results suggest that conversational implicatures based on violations of the quantity maxim are not always associated with a processing cost on the part of the hearer. The presence of conversational implicature is thus advantageous to the speaker without troubling the hearer, assuming that both are coordinated about the goal and direction of the conversation. If the latter condition is not met, the hearer's beliefs about the extent to which the speaker is cooperative and what she considers relevant can be mistaken, thus making room for notions like face (Pinker et al., 2008) and manipulation (Reboul, 2007). However, it is not necessary to evoke these notions to explain the emergence of conversational implicature. It is possible that the notions of face and manipulation coopted conversational implicature once it came into place through constraints of least effort.

6.4. Future research

A central assumption in this and much previous work on scalar inferences is that findings for "some" can be generalized to the entire family of scalar expressions. Indeed, the lexical access hypothesis predicts that there should be no processing difference between the inference from "some" to the negation of "all" and the inference from "warm" to the negation of "hot," since the derivation of both inferences involves substituting constituents with material from the lexicon.

However, recent work has shown that the assumption that all scalar expressions are alike is not always warranted (van Tiel, van Miltenburg, Zevakhina, & Geurts, 2014). The likelihood of scalar inferences varies dramatically for different scalar expressions, ranging from close to 0% to 100% with "some" occurring at the top of that scale. It thus seems at least conceivable that different scalar expressions might be associated with different processing profiles. Such processing differences might be related to differences in the number and availability of potential alternatives. For example, the space of possible alternatives for "That movie is good" is much larger than for "The battery is low" (see van Tiel et al., 2014: experiment 3). Hence, one might expect the inference from "good" to "not excellent" to be more time consuming than the inference from "low" to "not depleted." Extending the scope of inquiry to other scalar expressions thus seems a straightforward way of testing the robustness and generality of our results and conclusions.

If the results are corroborated, it will be interesting to see if children are less prone to derive scalar inferences for the same reason as adults who derive scalar inferences exhibit an increase in response times. To this end, the scope of inquiry in the developmental literature should be extended to other varieties of quantity implicature, such as free choice inferences (Tieu et al., 2015), conditional perfection, and exhaustivity in "it"-clefts (Heizmann, 2007). Given children's sensitivity to the particularities of the experimental task (Chierchia et al., 2001; Feeney et al., 2004; Guasti et al., 2005; Katsos & Bishop, 2010; Papafragou & Musolino, 2003), it will be important to test these inferences in an integrated experimental design.

Lastly, the lexical-access hypothesis predicts that it should be possible to eliminate the processing cost associated with the derivation of scalar inferences by priming the alternative expression, thus easing its retrieval and hence the construction of the required alternative. Moreover, assuming that children are less prone to derive scalar inferences because of their problems with retrieving alternatives from the lexicon, the same method should make their behavior more adultlike. Indeed, both predictions have been borne out in a number of experiments (Chierchia et al., 2001; Grodner et al., 2010; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004).

Regardless of the results of these investigations, the conclusion remains that many conversational implicatures are not associated with a processing cost. Only when their derivation involves competition with an alternative whose construction requires consulting the lexicon does an increase in reaction times occur. This finding warns us against using presence of a processing cost as a diagnostic for conversational implicatures. More broadly, our findings indicate that sometimes pragmatic inferencing is not associated with a processing cost.

Acknowledgments

This research was supported by F.R.S.-FNRS Research Incentive Grant F.4502.15 (principal investigator: Mikhail Kissine), the Bundesministerium für Bildung und Forschung (BMBF) Grant Nr. 01UG1411., and by the Research Council of the University of Leuven, project OT/11/030, which are gratefully acknowledged. We thank Corien Bary, Lewis Bott, Emmanuel Chemla, Chris Cummins, Michael Franke, Bart Geurts, Helen de Hoop, Paula Rubio-Fernández, and Sammie Tarenskeen for their comments on an earlier version of this paper, and the audiences of the Semantics and Pragmatics workshop at ZAS Berlin and the 2015 Experimental Pragmatics conference in Chicago for their valuable feedback. In addition, we thank the editor and reviewers of *Cognitive Science*. Their comments have made a tremendous impact on this paper and on our thoughts on the topic.

Notes

1. Mechanical Turk is a website where workers perform so-called human intelligence tasks (HITs) for financial compensation. It has been shown that the quality of data gathered through Mechanical Turk equals that of laboratory data (e.g., Buhrmester, Kwang, & Gosling, 2011; Schnoebelen & Kuperman, 2010; Sprouse, 2011).
2. The full experiment can be accessed online via http://spellout.net/ibexexps/bobvantiel/quantity-english1/experiment.html.

References

Atlas, J., & Levinson, S. (1981). It-clefts, informativeness, and logical form. In P. Cole (Ed.), *Radical pragmatics* (pp. 1–61). New York: Academic Press.

van der Auwera, J. (1997). Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics*, *27*(3), 261–274. doi:10.1016/s0378-2166(96)00058-6

Barker, C. (2010). Free choice permission as resource-sensitive reasoning. *Semantics and Pragmatics*, *3*(10), 1–38. doi:10.3765/sp.3.10

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93. doi:10.1016/j.cognition.2010.10.010

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 1.10-10. Available at http://CRAN.R-project.org/package=lme4. Accessed on January 27, 2016.

Bolinger, D. (1972). A look at equations and cleft sentences. In E. Firchow (Ed.), *Studies for Einar Haugen* (pp. 96–114). The Hague: Mouton de Gruyter.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. doi:10.1016/j.jml.2004.05.006

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. doi:10.1016/j.jml.2011.09.005

Breheny, R. (2006). Communication and folk psychology. *Mind and Language*, *100*(3), 434–463. doi:10.1111/j.1468-0017.2006.00307.x

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. doi:10.1177/1745691610393980

Büring, D., & Križ, M. (2013). It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics*, *6*(6), 1–29. doi:10.3765/sp.6.6

Byram-Washburn, M. (2012). The exhaustivity of it-clefts as a conversational implicature. Unpublished manuscript, University of Southern California.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45–73. doi:10.1037/h0076248

Chemla, E. (2009). Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics*, *2*(2), 1–33. doi:10.3765/sp.2.2

Chemla, E., & Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition*, *130*(3), 380–396. doi:10.1016/j.cognition.2013.11.013

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1741–1760. doi:10.1080/17470210701712960

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). Oxford, UK: Oxford University Press.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality of language." *Linguistic Inquiry*, *37*(4), 535–590. doi:10.1162/ling.2006.37.4.535

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domingues, & A. Johansen (Eds.), *Proceedings of the 25th Boston University conference on language development* (pp. 157–168). Somerville, MA: Cascadilla Press.

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *An international handbook of natural language meaning* (pp. 2297–2332). Berlin: Mouton de Gruyter.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology*, *11*, 430–477. doi:10.1016/0010-0285(79)90020-3

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517. doi:10.1016/0010-0285(72)90019-9

Coltheart, M. (2003). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). New York: Academic Press.

Crnič, L., Chemla, E., & Fox, D. (2015). Scalar implicatures of embedded disjunction. *Natural Language Semantics*, *23*(4), 271–305. doi:10.1007/s11050-015-9116-x

De Neys W., & Schaeken W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133. doi:10.1027/1618-3169.54.2.128

Deamer, F., Pouscoulous, N., & Breheny, R. (2010). A contrastive look at metaphor and hyperbole. *UCL Working Papers in Linguistics*, *22*, 1–16.

DeClerck, R. (1988). *Studies on copular sentences, clefts and pseudoclefts*. Leuven: Leuven University Press.

Degen, J. (2015). Investigating the distribution of "some" (but not "all") implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55. doi:10.3765/sp.8.11

Degen, J., & Tanenhaus, M. K. (2011). Making inferences: The case of scalar implicature processing. In L. Carlson, C. Höolscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3299–3304). Austin, TX: Cognitive Science Society.

Degen, J., & Tanenhaus, M. K. (2015a). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201.

Degen, J., & Tanenhaus, M. K. (2015b). Processing scalar implicature: A constraintbased approach. *Cognitive Science*, *39*(4), 667–710. doi:10.1111/cogs.12171

Destruel, E., & Farmer, T. A. (2015). Exhaustive inferences in on-line language comprehension. Poster presented at Experimental Pragmatics, Chicago.

DeVeaugh-Geiss, J. P., Zimmermann, M., Onea, E., & Boell, A.-C. (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. Poster presented at Experimental Pragmatics, Chicago.

Drenhaus, H., Zimmermann, M., & Vasishth, S. (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics*, *24*(3), 320–337. doi:10.1016/j.jneuroling.2010.10.004

Dufter, A. (2009). Clefting and discourse organization: Comparing Germanic and Romance. In A. Dufter & D. Jacob (Eds.), *Focus and background in Romance languages* (pp. 83–121). Amsterdam: John Benjamins.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*(2), 121–132. doi:10.1037/h0085792

von Fintel, K. (2001). Conditional strengthening: A case study in implicature. Unpublished manuscript, MIT. Available at http://web.mit.edu/fintel/fintel-2001-condstrength.pdf.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Houndmills: Palgrave Macmillan.

Fox, D., & Katzir, R. (2010). On the characterization of alternatives. *Natural Language Semantics*, *19*(1), 87–107. doi:10.1007/s11050-010-9065-3

Franke, M. (2009). Signal to act: Game theory in pragmatics. Ph.D. thesis, University of Amsterdam.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*(1), 1–82. doi:10.3765/sp.4.1

Franke, M. (2014). Typical use of quantifiers: A probabilistic speaker model. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the Cognitive Science Society* (pp. 487–492). Austin, TX: Cognitive Science Society.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. NewYork: Academic Press.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry*, *2*(4), 61–66.

Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics*, *13*(4), 383–410. doi:10.1007/s11050-005-2052-4

Geurts, B. (2010). *Quantity implicatures*. Cambridge, UK: Cambridge University Press.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures? *Semantics and Pragmatics*, *2*(4), 1–34. doi:10.3765/sp.2.4

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, *28*(4), 446–469. doi:10.1111/rati.12113

Geurts, B., & van Tiel, B. (2013). Embedded scalars. *Semantics and Pragmatics*, *6*(9), 1–37. doi:10.3765/sp.6.9

Gibbs, R. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 524–533. doi:10.1037//0278-7393.9.3.524

Gibbs, R. (1986a). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*, 3–25. doi:10.1037/0096-3445.115.1.3

Gibbs, R. (1986b). Skating on thin ice: Literal meaning and understanding idioms in conversation. *Discourse Processes*, *9*, 17–30. doi:10.1080/01638538609544629

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, *21*, 512–521. doi:10.1016/s0022-5371(83)90355-9

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, *21*, 85–98. doi:10.1016/s0022-5371(82)90467-4

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. doi:10.1016/j.cognition.2010.03.014

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696. doi:10.1080/01690960444000250

Halvorsen, P.-K. (1978). The syntax and semantics of cleft constructions. Ph.D. thesis, University of Texas at Austin.

Harris, R. (1976). Comprehension of metaphors: A test of the two-stage processing model. *Bulletin of the Psychonomics Society*, *8*, 312–314. doi:10.3758/bf03335150

Hartmann, K., & Veenstra, T. (2013). Introduction. In K. Hartmann & T. Veenstra (Eds.), *Cleft structures* (pp. 1–34). Amsterdam: John Benjamins.

Hauser, M. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hedberg, N. (2000). The referential status of clefts. *Language*, *76*(4), 892–920. doi:10.2307/417203

Heizmann, T. (2007). Child acquisition of exhaustivity in clefts. In H. Caunt-Nulton S. Kulatilake, & I. Woo (Eds.), *Proceedings of the 31st Annual Boston University conference on language development* (pp. 298–309). Sommerville, MA: Cascadilla.

Horn, L. R. (1972). On the semantic properties of logical operators in English. Ph.D. thesis, University of California, Los Angeles. Distributed by Indiana University Linguistics Club.

Horn, L. R. (1981). Exhaustiveness and the semantics of clefts. *Proceedings of the North East Linguistic Society (NELS)*, *11*, 125–142.

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, D.C.: Georgetown University Press.

Horn, L. R. (2000). From *if* to *iff*: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, *32*, 289–326. doi:10.1016/s0378-2166(99)00053-3

Horn, L. R. (2004). Implicature. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 2–28). Malden, MA: Blackwell.

Horn, L. R., & Bayer, S. (1984). Short-circuited implicature: A negative contribution. *Linguistics and Philosophy*, *7*(4), 397–414. doi:10.1007/bf00631074

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. doi:10.1002/bimj.200810425

Huang, Y. T., & Snedeker, J. (2009a). On-line interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. doi:10.1016/j.cogpsych.2008.09.001

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. doi:10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172. doi:10.1080/01690965.2010.508641

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*, 77–96. doi:10.1207/s15473341lld0202_1

Jennings, R. E. (1994). The “or” of free choice permission. *Topoi*, *13*(1), 3–10. doi:10.1007/bf00763643

Kamp, H. (1973). Free choice permission. *Proceedings of the Aristotelian Society*, *74*, 57–74. doi:10.1163/9789004252882_008

Katsos, N., & Bishop, D. V. M. (2010). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *20*(1), 67–81. doi:10.1016/j.cognition.2011.02.015

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, *30*(6), 669–690. doi:10.1007/s10988-008-9029-y

Kiss, E. (1998). Identification focus versus information focus. *Language*, *74*(2), 245–273. doi:10.1353/lan.1998.0211

Klinedinst, N. (2007). Plurals, possibilities, and conjunctive disjunction. *UCL Working Papers in Linguistics*, *19*, 261–284.

Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Y. Otsu (Ed.), *The proceedings of the third Tokyo conference on psycholinguistics* (pp. 1–25). Tokyo: Hituzi Syobo.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojensen Christensen, R. (2013). lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). Available at http://cran.r-project.org/package=lmerTest. Accessed January 27, 2016.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lilje, G. W. (1972). Uninvited inferences. *Linguistic Inquiry*, *3*, 540–542.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W.H. Freeman.

Martinet, A. (1962). *A functional view of language*. Oxford, UK: Clarendon Press.

Merin, A. (1992). Permission sentences stand in the way of Boolean and other lattice-theoretic semantices. *Journal of Semantics*, *9*(2), 95–162. doi:10.1093/jos/9.2.95

Millikan, R. (2005). *Language: A biological model*. Oxford, UK: Oxford University Press.

Morgan, J. (1978). Two types of convention in indirect speech acts. In P. Cole (Ed.), *Syntax and semantics, volume 9: Pragmatics* (pp. 261–281). New York: Academic Press.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. doi:10.1016/s0010-0277(00)00114-1

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210. doi:10.1016/s0093-934x(03)00053-1

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of “scalar inferences.” In N. Burton-Roberts (Ed.), *Advances in pragmatics* (pp. 184–212). Basingstoke, UK: Palgrave.

Onea, E., & Beaver, D. (2011). Hungarian focus is not exhausted. In E. Cormany, S. Ito, & D. Lutz (Eds.), *Proceedings of semantics and linguistic theory (SALT) 19*, pp. 342–359).

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language and meta-cognition* (pp. 140–169). Cambridge, UK: Cambridge University Press.

Ortony, A., Schallert, D., Reynolds, R., & Antos, S. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*, 465–477. doi:10.1016/s0022-5371(78) 90283-9

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *78*(3), 253–282. doi:10.1016/s0010-0277(02)00179-8

Papafragou, A., & Tantalou, N. (2004). Children’s computation of implicatures. *Language Acquisition*, *12*(1), 71–82. doi:10.1207/s15327817la1201_3

Pavey, E. L. (2004). The English it-cleft construction: A role and reference grammar analysis. Ph.D. thesis, University of Sussex.

Percus, O. (1997). Prying open the cleft. *Proceedings of the North East Linguistic Society (NELS)*, *27*, 337–351.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *24*(2), 169–190. doi:10.1017/s0140525x04000056

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13* (4), 707–784. doi:10.1017/S0140525X00081061

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Science*, *105*(3), 833–838. doi:10.1073/pnas.0707192105

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347–375. doi:10.1080/10489220701600457

Proust, J. (2002). Can 'radical' simulation theories explain psychological concept acquisition? In J. Dokic & J. Proust (Eds.), *Simulation and knowledge of action* (pp. 201–228). Amsterdam: John Benjamins.

R Development Core Team (2006). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reboul, A. (2007). *Langage et cognition humaine*. Grenoble: Presses Universitaires de Grenoble.

Récanati, F. (1995). The alleged priority of literal interpretation. *Cognitive Science*, *19*(2), 207–232. doi:10.1207/s15516709cog1902_2

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, *13*(1), 73–90. doi:10.1016/s0885-2014(98)90021-1

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116. doi:10.1007/bf02342617

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, *43*(4), 441–464. doi:10.2298/psi1004441s

Schulz, K. (2005). A pragmatic solution to the paradox of free choice permission. *Synthese*, *147*(2), 343–377. doi:10.1007/s11229-005-1353-y

Searle, J. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, volume 3: Speech acts* (pp. 59–82). New York: Academic Press.

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 83–111). Cambridge, UK: Cambridge University Press.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, *275*(5306), 1599–1603. doi:10.1126/science.275.5306.1599

Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry*, *13*(3), 483–545.

Sperber, D., & Wilson, D. (1987). Précis of relevance: Communication and cognition. *Behavioral and Brain Sciences*, *10*(4), 697–754. doi:10.1017/s0140525x00055345

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, *43*(1), 155–167. doi:10.3758/s13428-010-0039-7

van Tiel, B. (2012). Universal free choice? In A. Aguilar Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16: Volume 2* (pp. 627–638). Cambridge, MA: MIT Working Papers in Linguistics.

van Tiel, B. (2014a). Embedded scalars and typicality. *Journal of Semantics*, *31*(2), 147–177. doi:10.1093/jos/fft002

van Tiel, B. (2014b). Quantity matters: Implicatures, typicality, and truth. Ph.D. thesis, Radboud University Nijmegen.

van Tiel, B., & Geurts, B. (2014). Truth and typicality in the interpretation of quantifiers. In U. Etxeberria, A. Fălăuş, A. Irurtzun, & B. Leferman (Eds.), *Proceedings of Sinn und Bedeutung 18* (pp. 451–468). Bayonne and Vitoria-Gasteiz.

van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar diversity. *Journal of Semantics*, *33*(1), 137–175. doi:10.1093/jos/ffu017.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2015). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*. doi:10.1093/jos/ffv001 In press.

Tomlinson Jr., J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*(1), 18–35. doi:10.1016/j.jml.2013.02.003

Vallduvì, E. (1993). The informational component. Ph.D. thesis, University of Pennsylvania.

Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 411–448). Oxford, UK: Oxford University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. doi:10.1016/0010-0277(83)90004-5

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, *8*(4), 255–290. doi:10.1023/A%3A1011255819284

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. The original experiment.